

# The strong side of weak topological insulators

Zohar Ringel,\* Yaacov E. Kraus,\* and Ady Stern

*Department of Condensed Matter Physics, Weizmann Institute of Science, Rehovot 76100, Israel*

Three-dimensional topological insulators are classified into “strong” (STI) and “weak” (WTI) according to the nature of their surface states. While the surface states of the STI are topologically protected from localization, this does not hold for the WTI. In this work we show that the surface states of the WTI are actually protected from any random perturbation that does not break time-reversal symmetry, and does not close the bulk energy gap. Consequently, the conductivity of metallic surfaces in the clean system remains finite even in the presence of strong disorder of this type. In the weak disorder limit the surfaces are found to be perfect metals, and strong surface disorder only acts to push the metallic surfaces inwards. We find that the WTI differs from the STI primarily in its anisotropy, and that the anisotropy is not a sign of its weakness but rather of its richness.

PACS numbers: 73.43.-f, 73.43.Cd, 73.20.-r, 71.23.-k, 72.15.Rn

## I. INTRODUCTION

Topological insulators (TIs) have recently become a very active subject in condensed matter physics. The classification of states of matter according to topological indices opens new horizons both theoretically and experimentally, and may hopefully lead to applications<sup>1,2</sup>.

For more than two decades, topological classification of phases was manifested primarily in the realm of the quantum Hall effect. The experimental observations of TI in two dimensions<sup>3</sup> (2D) and three dimensions<sup>4</sup> (3D) expanded this notion also to systems that are time-reversal (TR) symmetric, and have sparked a “race for golf” for new topological phases, and for their unique properties.

In 2D, TR symmetric band insulators are classified into “trivial” and “topological” by a  $\mathbb{Z}_2$  index<sup>5</sup>. At the one-dimensional (1D) interfaces between a topological insulator and the vacuum (or any other trivial insulator), the energy gap must close, implying the appearance of counter-propagating chiral gapless modes. As long as TR symmetry is preserved, these modes are protected from back-scattering and gapping. In contrast, when a bi-layer system is formed of two such TIs, coupling of the edge modes in the two layers may gap them without violating the TR symmetry.

In 3D, TI's are classified by four  $\mathbb{Z}_2$ -indices  $(\nu_0, \boldsymbol{\nu})$ <sup>6–9</sup>. A non-trivial  $\nu_0$  implies that on each 2D surface of the sample, the bulk gap is closed by surface states, the spectrum of which consists of an odd number of Dirac cones. As long as TR symmetry is preserved and the bulk gap remains open, at least one Dirac cone will survive the addition of any perturbation. Moreover, the wavefunctions of this Dirac cone can not be localized by disorder, and the surface of the 3D TI is apparently a perfect metal in the absence of electron-electron interaction<sup>10–13</sup>. Because of the robustness of its surface states, this phase was called “strong TI” (STI)<sup>6</sup>.

On the other hand, if  $\nu_0 = 0$  but  $\boldsymbol{\nu} \neq 0$ , the system is in a phase known as a “weak TI” (WTI). This phase is adiabatically connected to stacked layers of 2D TI's<sup>6</sup>. Suppose we have a cubic sample. The two sur-

faces which are aligned with the top and bottom layers will in general be gapped. But, the four perpendicular surfaces have gapless states, at least in a clean system. In the limit of completely decoupled layers, these surface states are actually the edge states of the stacked 2D TI. Translation-invariant coupling between the layers gaps out most of these surface states. However, Kramer's theorem ensures two Dirac cones to remain, both centered at momenta that are TR invariant. In the following, we refer to this type of surfaces, unless otherwise stated.

The chief reason why the WTI is considered weak is that its surface modes may be gapped without breaking TR symmetry or closing the bulk gap. In the stacked-layers picture, a mass term that gaps the edge modes arises if one couples the layers in pairs. The only symmetry violated by this term is the lattice-translation symmetry. Therefore, it appears that this symmetry is essential for the topological protection of the WTI surfaces. Since disorder breaks translational symmetry, one may be led to assume that the WTI surfaces are no longer protected and behave like conventional 2D metals with strong spin-orbit couplings. Such metals are known to undergo an Anderson transition from metals to insulators as a function of disorder strength<sup>14</sup>.

In this paper, we show that the contrary is true. We consider the effect of disorder on the weak TI, and show that it is actually not weak at all. In Sec. II, we show that the conductivity of the non-trivial surfaces of the WTI remains higher than  $e^2/h$  in the presence of disorder of arbitrary strength, as long as the bulk gap and TR symmetry are maintained.

Section III includes perturbative analysis. In the limit of weak disorder, we evaluate the weak localization correction, and find it to be anti-localizing. In the opposite limit, we consider strong disorder that is limited to several atomic layers at the surface of the insulator. We find that such disorder makes the surface insulating, but creates a perfect metallic sheet just beneath the disordered surface. We also discuss the conductivity in the intermediate disorder limit, and raise the possibility that a phase with a universal finite conductivity appears.

In light of these results, we discuss in Sec. IV the unique surface anisotropy of the WTI, which implies that the robustness of the conductivity of a surface strongly depends on its orientation. Based on this anisotropy we raise the possibility of surface engineering.

## II. FINITE CONDUCTIVITY

Assume one stacks an even number of 2D TI's, and couples them in pairs. Each such pair is topologically trivial, and generically has an insulating edge. Thus, in the 3D limit of an infinite even number of layers the surfaces are generically insulating. On the other hand, if the number of layers is odd, there is no way to gap all the edge modes without breaking TR symmetry, and the surface must be conducting. This sensitivity to the parity of the layer number was then argued to imply the fragility of the WTI<sup>15</sup>.

While the argument for non-triviality in the odd case relies on topology, the argument for gapping of the surface modes in the even case relies on a well-tailored perturbation that couples the layers in pairs. Random disorder does not induce such a coherent perturbation. Rather, when disorder is present and the number of layers is even, the surfaces *may* be trivial, yet do not have to be. On the other hand, for an odd number of layers, the surfaces *must* conduct. This suggests that when the coupling between layers is disordered, the odd behavior is in fact the generic one, thus the surfaces will conduct for any large number of layers.

This heuristic argument will now be put on firm theoretical ground. Consider a WTI of dimension  $L^3$  which is adiabatically connected to an odd number of 2D layers stacked along the  $\hat{z}$  direction and  $L \gg 1$ . Note that we take the lattice spacing to be 1. We take the periodic boundary conditions to be periodic in the  $\hat{z}$  and  $\hat{x}$  directions, and open in the  $\hat{y}$  direction, as illustrated in Fig. 1(a). Under these boundary conditions, the surface states reside on the interior and exterior surfaces of a thickened torus. We allow for any disorder which is TR symmetric, does not close the bulk gap, and has a correlation length much smaller than  $L$ . Under these conditions, the surfaces have no special regions or lines to which the electrons wave functions could be restricted.

Consider an Aharonov-Bohm flux that implements a phase twist  $\phi$  in the periodic boundary conditions along the  $\hat{x}$  direction, as illustrated in Fig. 1(a). Let us study how the spectrum of the edge modes depend on  $\phi$ . For  $\phi = 0, \pi$ , the Hamiltonian is TR symmetric, and Kramer's theorem guarantees that all the energies are doubly degenerate. Apart from these degeneracies, the spectrum has no accidental degeneracies, as implied by the non-crossing theorem<sup>16</sup>. This ensures a well-defined labeling of energies as a function of  $\phi$ ,  $E_i(\phi)$ , where  $i = 1, 2, \dots$  and  $E_{i+1} \geq E_i$ .

The difference between topologically trivial and non-trivial surfaces is manifested in the relation between the

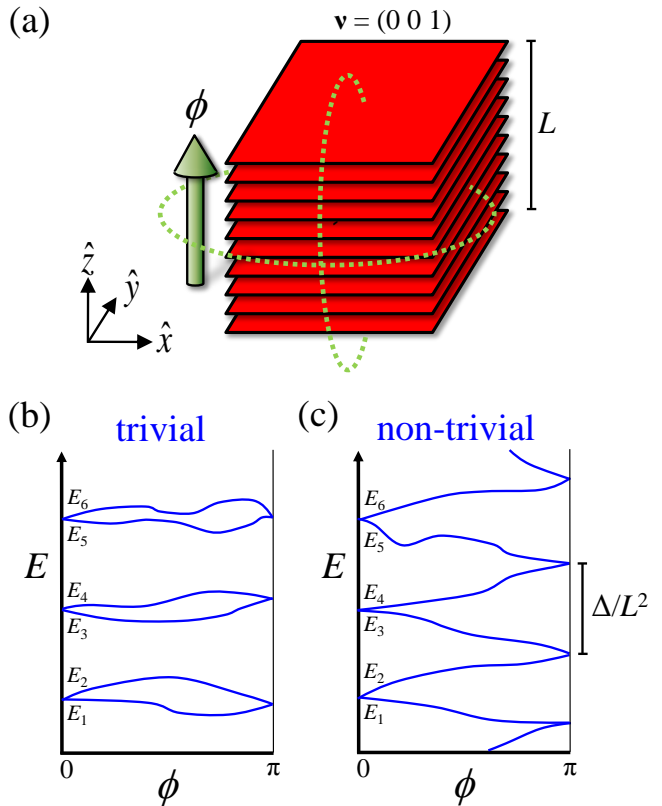


FIG. 1: Topologically trivial and non-trivial pair switching. (a) The WTI is adiabatically connected to stacked layers of 2D TI. We consider a WTI of dimension  $L^3$  with  $\mathbf{v} = (0 0 1)$  and  $\hat{z}$  as the stacking direction. The boundary conditions (green dotted lines) are periodic in the  $\hat{z}$  direction and are twisted by an Aharonov-Bohm flux in the  $\hat{x}$  direction (green thick arrow). The remaining  $xz$  surfaces are metallic. (b), (c) Typical patterns of energies of surface state as a function of the Aharonov-Bohm flux,  $\phi$ , for (b) trivial and (c) non-trivial surfaces. Kramer's theorem assures that at the time-reversal invariant fluxes  $\phi = 0, \pi$ , states come in degenerate pairs. On a trivial surface, the pairs remain the same between these two values, while on a non-trivial surface the pair switch partners. The mean level spacing of the surface states is  $\Delta/L^2$ , where  $\Delta$  is the bulk gap. We show that non-trivial pair switching implies the existence of at least  $O(L)$  extended states.

pairs of degenerate states at  $\phi = 0$  and at  $\pi$ <sup>6,17,18</sup>, which is illustrated in Fig. 1(b) and 1(c). If the pairs at  $\phi = 0$  are the same as those at  $\phi = \pi$ , the surface is topologically trivial. This is the case for a trivial insulator and for a WTI with an even number of 2D layers. In contrast, if the pairs switch partners between  $\phi = 0$  and  $\pi$ , the surface is non-trivial. Such pair switching takes place on the surfaces of an STI and of a WTI with an odd number of 2D layers. The zigzag shape of the spectrum in the non-trivial surfaces cannot be terminated without

approaching the bulk states. Hence, in this case

$$\sum_i |E_i(\pi) - E_i(0)| \geq \Delta, \quad (1)$$

where the summation is over all surface states and  $\Delta$  is the bulk gap. Note that for any finite  $L$ , the number of surface states is proportional to  $L^2$ , and the mean level spacing is  $\Delta/L^2$ .

It is impossible to satisfy inequality (1) if all the surface states are exponentially localized. The current of a localized state is exponentially small with the system size. The current carried by an electron in the  $i^{\text{th}}$  eigenstate is given by  $I_i(\phi) = (e/h)\partial_\phi E_i$ <sup>19</sup>. Therefore,  $\partial_\phi E_i \sim e^{-L}$ , and consequently,  $|E_i(\pi) - E_i(0)| \sim e^{-L}$ . In that case inequality (1) cannot be satisfied.

Furthermore,  $I_i = e\langle v \rangle_i/L$ , where  $\langle v \rangle_i$  is the expectation value of the velocity. Since, the velocity is bounded by intrinsic variables and cannot increase with the system size,  $I_i$  approaches zero at least as  $1/L$ . Note that there is a value  $\phi_0$  such that  $|E_i(\pi) - E_i(0)| \leq (\pi h/e)|I_i(\phi_0)|$ . Therefore, for the inequality (1) to be satisfied, there must be at least  $O(L)$  delocalized states. Furthermore, as long as the system is homogeneous, which is the case for random disorder, these states are distributed all over the surface.

Imagine now cutting the system into two subsystems, one with even and one with odd number of layers. Since the cut is a surface effect and the system is homogeneous, it will not localize a state that has been delocalized before. Thus, in the presence of random disorder, delocalized states will exist also in the subsystem with the even number of layers. We can therefore conclude that in a system with an even number of layers there are delocalized states in the presence of random disorder, despite the absence of topological protection.

The homogeneity of the disordered system leads to a further consequence. Suppose we have cubic slabs of dimension  $l^3$ . According to what we have seen before, on the surface of the small cubes there are at least  $O(l)$  delocalized states on the scale of  $l$ . Now we glue the cubes to one another and obtain a larger cube of dimension  $L^3$ . Since the gluing process does not localize states, on the surface of the large cube there are at least  $O(L^2/l^2)$  delocalized states on the scale of  $l$ . This scaling is consistent both with delocalization of all states, and with a scenario of localized states with a broad distribution of localization lengths.

Finally, we notice that for the current  $I_i$  to decay as  $1/L$ , the electronic motion must be ballistic. If, however, the motion is diffusive, then the current decays as  $1/L^2$ . In ballistic motion, inequality (1) required  $O(L)$  delocalized states. In contrast, a diffusive motion requires  $O(L^2)$  such states. Since ballistic motion is unlikely in the presence of disorder, the bound of  $O(L)$  states is probably too restrictive.

Having showed the existence of delocalized states, we can turn to estimate a lower bound for the conductivity.

We use the Thouless formula, which relates the electrical conductivity to the sensitivity of energies to phase twists<sup>20–23</sup>,

$$\sigma_{xx} \approx \frac{e^2}{h} \left\langle \frac{\Delta E}{\Delta \phi} \right\rangle \frac{dN}{dE}, \quad (2)$$

where  $\langle \Delta E/\Delta \phi \rangle$  denotes geometric mean of the energy difference  $E_i(\pi) - E_i(0)$  averaged over eigenstates and  $dN/dE$  denotes the density of states, both at Fermi energy. This relation has been shown to be only qualitatively correct<sup>24</sup>. For example, in 1D systems the conductivity scales like  $[\langle \Delta E/\Delta \phi \rangle (dN/dE)]$ <sup>225</sup>, and constants of order unity may appear<sup>22,26</sup>. Moreover, discrepancies of  $O(1)$  may appear if the relation is expressed with  $\partial^2 E/(\partial \phi)^2$ , rather than with  $\Delta E/\Delta \phi$ <sup>22</sup>. Nevertheless, when  $\langle \Delta E/\Delta \phi \rangle (dN/dE)$  is of the order of unity, the conductivity is expected to be of the order of  $e^2/h$ .

In the non-trivial pair switching, see Fig. 1(c), the zigzag shape of the spectrum relates  $\langle \Delta E/\Delta \phi \rangle$  to the energy levels spacing  $E_{i+1} - E_i$ . And since the level spacing is the inverse density of states, it leads to  $\langle \Delta E/\Delta \phi \rangle (dN/dE) \geq 1$ . Consequently,

$$\sigma_{xx} \geq \frac{e^2}{h}. \quad (3)$$

We have therefore arrived at our key result: a non-trivial surface of a WTI will remain conducting even in the presence of random disorder.

Preliminary numerical work, reported in Appendix A, indeed shows that as the number of stacked layers increases, the even-odd difference diminishes, and both tend to lack of localization.

### III. PERTURBATIVE ANALYSIS

The topological argument allowed us only to bound the conductivity. More quantitative predictions can be given in the limits of weak disorder and strong surface disorder, where perturbative approaches can be utilized. Disorder is defined to be weak when  $E_F \tau \gg 1$ , where  $E_F$  is the Fermi energy and  $\tau$  is the mean free time. In this limit we evaluate the lowest-order quantum correction to the conductivity<sup>27</sup>.

The low energy effective Hamiltonian describing the surface of a WTI in the clean limit consists of two decoupled Dirac cones

$$H(k_x, k_y) = v_0(k_x I^* \otimes s_x + k_y I \otimes s_y). \quad (4)$$

For every value of  $k_x$  and  $k_y$  this is a  $4 \times 4$  matrix, spanned by a direct product of two Pauli spinors:  $\tau$  that denotes the Dirac cone, and  $\mathbf{s}$  that denotes the electron spin. Here  $v_0$  is the velocity characterizing the Dirac cones, and  $I^*$  may be either the unity matrix  $I$  or the Pauli matrix  $\tau_z$ , depending on the particular WTI considered. The corresponding TR operator is  $T_W = I \otimes i s_y K$ , where  $K$

denotes complex conjugation. Accordingly,  $T_W^2 = -1$ . Notably, since under TR each Dirac cones is mapped to itself, in general their chiralities are unrelated, as well as the energy of the Dirac points.

Disorder adds to the Hamiltonian a sum of the form  $\sum_{m,n} V_{mn}(\mathbf{r})(\tau_m \otimes s_n)$ , where the indices  $m, n$  take the values  $0, x, y, z$  and  $\tau_0 = s_0 = I$ . For the WTI only six terms are TR symmetric:  $I \otimes I, \tau_z \otimes I, \tau_x \otimes I$  and  $\tau_y \otimes \mathbf{s}$ . The first three describe potential disorder, and the last three describe random spin-orbit scattering (note that the clean Hamiltonian already includes spin-orbit). Among the six, only the term  $\tau_y \otimes s_z$  gaps the spectrum.

In Appendix B we evaluate the lowest order quantum correction to the conductivity from the low energy Hamiltonian (4), in the presence of all mentioned types of disorder<sup>28</sup>. We find this correction of be anti-localizing,

$$\frac{d \ln \tilde{\sigma}_{xx}}{d \ln L} = -T_W^2 \frac{1}{2\pi \tilde{\sigma}_{xx}} f_v^2 \left(1 - \frac{f_e}{2}\right) > 0, \quad (5)$$

where  $\tilde{\sigma}_{xx} = \sigma_{xx}(h/e^2) = E_F \tau f_v$ . Furthermore,  $2/3 < f_v < 2$  is the vertex correction, and  $-1 \leq f_e \leq 1$  is a correction of the Cooperon, both are determined by the details of the disorder. Equation (5) implies that the conductivity flows towards a perfect metal, and  $\tilde{\sigma}_{xx}$  increases logarithmically with the system size.

The Hamiltonian (4) appears also in two other 2D systems, and it is instructive to elucidate the similarities and differences between these systems and the WTI. The first system is that of spinless electrons in graphene. For that system  $I^* = \tau_z$ , and  $\mathbf{s}$  denotes the sublattice index. Accordingly, its TR operator is  $T_G = \tau_x \otimes IK$  and  $T_G^2 = 1$ . By plugging  $T_G$  instead of  $T_W$  into Eq. (5), we observe in graphene weak localization, as expected<sup>29,30</sup>. As a matter of fact, since for the WTI  $T_W^2 = -1$ , for generic disorder the Hamiltonian belongs to the symplectic class, which is known to have weak anti-localization correction<sup>14</sup>. In contrast, spinless graphene, for which  $T_G^2 = 1$ , belongs to the orthogonal class, which shows weak localization. The general relation between the symmetry class and the sign of the weak-localization correction can be shown using the non-linear  $\sigma$  model approach<sup>14</sup>, but may be also understood more directly by means of interference of diffusive trajectories, as shown in Appendix C.

The second system is that of a 2D insulator at the transition point between a trivial and a topological phase in the absence of inversion symmetry<sup>31</sup>. This system belongs to the same symmetry class as the WTI, but its spectrum does not exhibit pair switching as a function of flux. Nonetheless, since it is tuned to a phase transition, one expects the correlation length to diverge and therefore delocalized states must exist at low energies. In Refs.<sup>32-34</sup> it was established that a band of delocalized states appears around zero energy, while far from zero energy states are localized. This should be compared with the WTI, where, as we have argued, states remain delocalized for any sub-gap energy. This discrepancy suggests that while these models have similar low

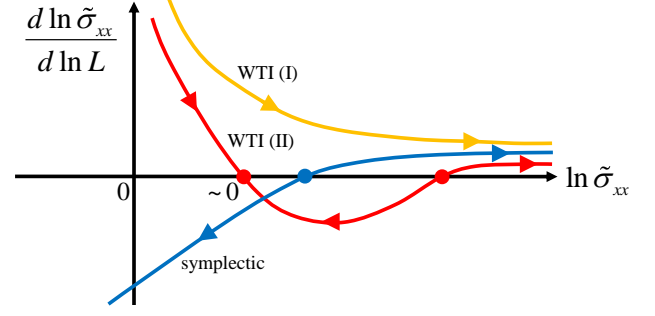


FIG. 2: Renormalization group flow of the conductivity. The  $\beta$ -function of the dimensionless conductivity  $\tilde{\sigma}_{xx}$  for a non-trivial surface of the WTI, compared with that of a 2D metal with strong spin-orbit coupling that belongs to the symplectic class (blue). According to Eq. (5) in the limit of high conductivity the flow is toward a perfect metal. We showed that the conductivity of a WTI surface cannot drop below  $e^2/h$ . Consequently, two types of flows are possible: (I) always flowing towards a perfect metal (yellow), and (II) flowing with a stable fixed point of finite conductivity (red).

energy descriptions, they are nonetheless different when the entire spectrum is taken into account.

Notably, following the posting of this manuscript on the arXiv, our prediction for delocalization within the low energy theory, Eq. (4), was validated numerically<sup>35</sup>. The restriction of this numerical work to low energy does not address directly the role of pair switching. Indeed, one can find a unitary transformation that maps the low energy Hamiltonian used there, including the disorder terms, to the low energy part of the Hamiltonian used in Ref.<sup>31</sup> to describe the transition mentioned in the previous paragraph, which exhibits no pair-switching.

Having shown that in the limit of weak disorder we have a perfect metallic surface, we now turn to opposite the limit of extremely strong disorder. In this limit the strength of the disorder is much larger than all the other energy scales, including the bulk band width and band gap. If such disorder acts on the entire 3D system, it mixes the bulk bands and makes the entire sample a trivial insulator. However, an interesting case is disorder that is limited to several of the outmost layers. This may actually happen in realistic surfaces, which are usually made dirty by oxides and other dopants. Moreover as we show below, it also reveals the role of the bulk in protecting the surface states.

Let us divide the Hamiltonian of the three dimensional system,  $H_{3D}$ , into the part that operates only within the clean bulk ( $H_0$ ), the part that operates on the disordered surface layers ( $H_{dis}$ ), and the part of hopping between the two ( $V$ ). The Hamiltonian may now be written as

$$H_{3D} = \begin{pmatrix} H_0 & V \\ V^\dagger & H_{dis} \end{pmatrix}. \quad (6)$$

We begin with the case where all the eigenvalues of  $H_{\text{dis}}$  are greater in absolute value than some value  $W$ , and  $W \gg t$ , where  $t$  is the bulk band width. For this case, all the eigenstates of  $H_{\text{dis}}$  are localized on the surface. These eigenstates may be considered as an high energy sector, which can be integrated out. To this end, we consider the Green's function projected onto the Hilbert space of the clean bulk<sup>36</sup>, using the projection operator  $P_0$

$$P_0(E - H_{3D})^{-1}P_0 = (E - H_0 - V(H_{\text{dis}} - E)^{-1}V^\dagger + O(V^4))^{-1}. \quad (7)$$

This Green's function defines an effective Hamiltonian for the clean bulk, which is

$$H_{\text{eff}}(E) = H_0 + V(H_{\text{dis}} - E)^{-1}V^\dagger + O(V^4). \quad (8)$$

This effective Hamiltonian describes the degrees of freedom of a 3D WTI which is clean in the bulk (the first term), and is disordered at its surface (the second term). Note that this surface lies beneath the physical surface, where the disorder vanishes. The second term of Eq. (8) represents virtual hopping from the bulk to the strongly localized states at the physical surface and back. Since all the eigenvalues of  $H_0$ , the matrix elements of  $V$  and the energy  $E$  are of the order of  $t$ , this bulk-surface coupling is of the order  $t^2/W \ll t$ . The effective Hamiltonian then describes a *weakly disordered* WTI, the gapless states of which are located underneath the physical surface. Recalling the above result, we can see that the relocated surface states form a perfect metal.

The same holds when the spectrum of  $H_{\text{dis}}$  becomes continuous. The states in the strongly disordered layers with energy greater than  $W$  can still be integrated out, resulting in small  $O(t^2/W)$  terms. The remaining low lying states are expected to be localized, since  $H_{\text{dis}}$  alone is not protected from localization. Such states act as strong scatterers. However, their density is of  $O(t/W)$ , and is therefore small, yielding a long mean-free path. Hence, we are still in the limit of weak disorder, thus having a perfect metal.

Intermediate disorder is disorder with  $E_F\tau \ll 1$  but  $\Delta\tau \gg 1$ . According to the topological argument, the conductivity has to be larger than  $e^2/h$  even in this regime. Following the single parameter scaling approach, two possible flows of the renormalization group may arise, which can be presented in terms of the  $\beta$ -function,  $\beta(\tilde{\sigma}_{xx}) = d \ln \tilde{\sigma}_{xx} / d \ln L$ . In one flow, the conductivity always flows to infinity while increasing the system size, as presumably happens in the STI<sup>10-13</sup>. In the second flow, a stable fixed point appears at  $\tilde{\sigma}_{xx} \approx 1$ , and a critical point appears for some  $\tilde{\sigma}_{xx} > 1$ . The two flows are illustrated in Fig. 2.<sup>37</sup>

#### IV. SURFACE ANISOTROPY

In the previous sections we analyzed the conduction properties of surfaces of the WTI, and found that they

are conducting even in the presence of disorder. This robustness brings the WTI closer to the STI in terms of their transport properties. Nevertheless, the WTI differs from the STI in the unique anisotropic behavior of its surfaces, which gives rise to the idea of surface engineering.

While non-trivial surfaces of the WTI are indeed robustly conducting, not all possible surfaces of the WTI are topologically non-trivial. For example, we mention that in the stacked-layers picture the top and bottom surfaces are topologically trivial, and are generally gapped. For given weak indices  $\nu$  and a plane with Miller indices  $\mathbf{h}$ , we define the relation  $\mathbf{h} \sim \nu$  by

$$(h_i - \nu_i) \mod 2 = 0, \quad (9)$$

for  $i = 1, 2, 3$ . Any surface with Miller indices that satisfy this relation is topologically trivial, whereas a surface with  $\mathbf{h} \approx \nu$  is topologically non-trivial<sup>6</sup>. The reason of this criterion is that the indices vector  $\nu$  does not uniquely define a stacking direction, and any vector  $\mathbf{h} \sim \nu$  can be a stacking direction, as illustrated in Fig. 3(a). Namely, the stacked-layers picture is a theoretical construction rather than a physical description, and in practice, the WTI does not have to be layered.

An alternative explanation for criterion (9) can be given from the picture of a fixed stacking direction and varying surfaces. Consider a WTI, the primitive lattice vectors of which are  $\mathbf{a}_1, \mathbf{a}_2, \mathbf{a}_3$ , making the lattice sites located at  $\mathbf{r}_n = \sum_{i=1}^3 n_i \mathbf{a}_i$ . Consider also a surface with Miller indices  $\mathbf{h}$ , and for simplicity place the origin of the coordinate system at some lattice site of the surface. By definition, all the lattice sites on the surface satisfy the condition  $\mathbf{h} \cdot \mathbf{n} = 0$ . For simplicity, we take the example of  $\nu = (001)$  and choose it to be the stacking direction. If  $\mathbf{h} \sim \nu$ , then  $h_3$  is odd, while  $h_1$  and  $h_2$  are even. Accordingly, on the surface all the  $n_3$  coordinates are even, and adjacent surface sites differ by an even increment of  $n_3$ . Therefore, the surface is composed of steps of an even number of layers, as illustrated in Fig. 3(a), and the coupling between them will gap the edge states. For  $\mathbf{h} \approx \nu$ , the surface is composed of steps of odd layers. Now, the coupling can not gap all the edge states, and the surface will conduct.

The high and non-trivial sensitivity of the surfaces to their orientation even in the presence of disorder is demonstrated in Fig. 3(b). We considered a  $20 \times 20 \times 20$  lattice of the Fu, Kane, and Mele model with  $\lambda_{SO} = t$  and  $\delta t = [-0.6, 0, 0.2, 0]t$ , which corresponds to  $\nu_0 = 0$  and  $\nu = (001)$  with a bulk gap of  $\Delta = 0.8t$ . In this model  $\nu$  represents also the weak hopping direction. Uniformly distributed strong disorder of magnitude  $t$  was also introduced. The figure depicts the local density of surface states integrated over an energy window  $|E| < 0.1\Delta$ . The parallelepiped is cut along the primitive vectors, and therefore has 2 trivial gapped faces and 4 topological metallic faces.

The criterion for a surface  $\mathbf{h} \sim \nu$  implies that the spectrum on it will be gapped, but it does not provide

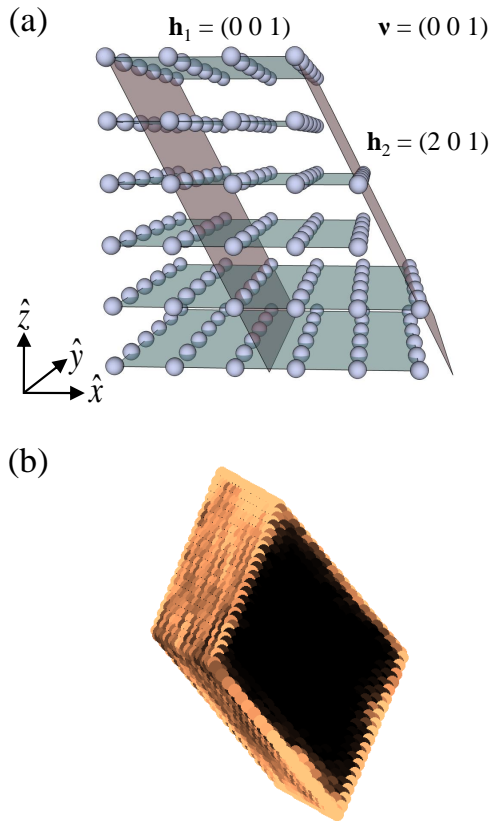


FIG. 3: Insulating and metallic surfaces. (a) The WTI has both trivial and non-trivial surfaces. A surface with Miller indices  $\mathbf{h}$  is trivial if  $(\mathbf{h} - \boldsymbol{\nu}) \bmod 2 = 0$ , denoted by  $\mathbf{h} \sim \boldsymbol{\nu}$ , since any such  $\mathbf{h}$  can denote a stacking direction. The figure depicts two trivial surfaces  $\mathbf{h}_1 = (001)$  and  $\mathbf{h}_2 = (201)$  for a cubic crystal with assumed  $\boldsymbol{\nu} = (001)$ . Both  $\mathbf{h}_1$  and  $\mathbf{h}_2$  are legitimate stacking directions. Alternatively, for a stacking along  $\mathbf{h}_1$ , the  $\mathbf{h}_2$  surface is composed of steps of two layers. The coupling between the layers gaps their edge states. For  $\mathbf{h} \sim \boldsymbol{\nu}$ , the steps will be of odd number of layers and will therefore conduct. (b) An example of the surface anisotropy in the Fu, Kane and Mele model of the weak  $\boldsymbol{\nu} = (001)$  phase. Depicted is the local density of surface states integrated over an energy window  $|E| < 0.1\Delta$ , with disorder strength comparable to the bulk gap. The surfaces of the parallelepiped are spanned by the primitive vectors. The two faces with Miller indices equal to  $\boldsymbol{\nu}$  are gapped, while the other four, with orthogonal Miller indices, are metallic. By controlling the cleavage process, the conductance of each face of the WTI can be engineered.

information on the magnitude of the gap. In the above example, for  $\mathbf{h}$  chosen to be in the weakest hopping direction, the energy gap on the surface is comparable to the bulk gap. Other trivial surfaces have energy gaps much smaller than this value. The influence of disorder on the gap and localization length of such surfaces may be dramatic. We note that for a surface that cannot be

described by Miller indices, we expect metallic behavior, since the scaling argument which was used to ensure  $\sigma_{xx} \geq e^2/h$  seems to hold.

By noticing that the topological and trivial surfaces are isotropically distributed, one can imagine creating a sample with each face engineered to be either gapped or metallic. A gapped surface along a stacking direction would remain insulating, while other surfaces will conduct. Provided rather good control on the cleaving process, various different electronic behaviors are expected on different surfaces, ranging all the way from perfect metals to insulators with varying gaps. In light of these results, we find that the anisotropic behavior of the WTI surfaces is not a sign of their weakness, but rather of their richness.

## V. SUMMARY

In this work, we showed that the name “weak topological insulators” does not do justice to the phase it describes, since the electrical conductivity of the non-trivial surfaces of such insulators is not suppressed by disorder. The WTI shows unique sensitivity of the electronic properties of its surfaces to their orientation, and that may provide an experimental tool for controlling these properties. We hope that this work will serve as a trigger for further study of these interesting topological phases.

## Acknowledgements

The authors thank Y Imry, FDM Haldane and IA Gruzberg for useful discussions. ZR thanks ISF grant 700822030182. YEK and AS thank the US-Israel Binational Science Foundation, the Minerva foundation and Microsoft’s station Q for financial support.

## Appendix A: Numerical analysis of disordered thin WTI

In an attempt to address numerically the effect of disorder on the conductivity of a gapless surface of the WTI, we considered the Fu, Kane and Mele model<sup>6</sup> with  $\lambda_{SO} = t$  and  $\delta t = [-0.6, 0, 0.2, 0]t$ , which corresponds to a  $\nu_0 = 0$  and  $\boldsymbol{\nu} = (001)$  with a bulk gap of  $\Delta = 0.8t$ . We also took the chemical potential to be at the Dirac points of the surface spectrum. The most general potential disorder that is symmetric to time-reversal was included by adding a time reversal symmetric random matrix which acts within unit cells. The entries of each matrix were sampled from a uniform distribution in some region  $[-w/2, w/2]$ , and the resulting matrix was then symmetrized with respect to time reversal. The disorder was added on three outmost layers with  $w = 0.5t, 0.5e^{-2}t, 0.5e^{-3}t$  corresponding to the first, second and third layer respectively. The samples sizes



$L_x \times L_y \times L_z$  ranged in from  $40 \times 10 \times 1$  up to  $120 \times 10 \times 6$  unit cells, where  $L_z$  can be thought as the number of the stacked 2D layers.

In order to obtain the conductance  $g_{xx}$  we used Eq. (2) of the main text. When applied to a quasi-1D sample this equation yields the conductance rather than the conductivity<sup>21</sup>. The fluctuations in energy levels following the insertion of a  $\pi$ -twist were approximated by extrapolating the derivative of the energy levels with respect to the phase twist. The geometric averaging was taken over the different instances of disorder and over an energy window of  $[-0.2t, 0.2t]$ . Although this second averaging is not included in the definition, we find that it did not have significant influence on the asymptotic behavior. We considered 30 instances of disorder for 1-3 layers, and 10 instance of disorder for 4-6 layers. The error bars are primarily due to fluctuations of the density of states which limit the accuracy of the estimated mean value.

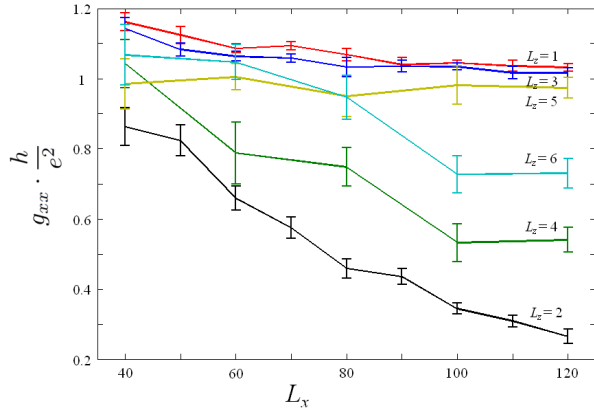


FIG. 4: The conductance  $g_{xx}$  for a WTI with  $\nu = (001)$  of the Fu, Kane and Mele model as estimated from the flux sensitivity of surface state energies multiplied by the density of states. Each line corresponds to a given number of layers ( $L_z$ ) and shows  $g_{xx}$  as a function of  $L_x$ , where  $L_y$  is fixed to 10. Periodic boundary conditions were imposed on  $\hat{z}$  and  $\hat{x}$ . Samples with an odd number of layers have a topologically protected minimal conductance of  $e^2/h$ . Samples with an even number of layers are (strictly speaking) topologically trivial and show a localization behavior. However their conductance converges to that of the odd layers as the number of channels is increased.

The dependence of  $g_{xx}$  on the dimensions of the surface is depicted in Fig. 4. For odd  $L_z$  the conductance tends to values close to the  $e^2/h$ , and shows no sign of localization. For comparison, the localization length of a sample with disorder of a similar strength that does not satisfy time reversal symmetry is around 40 unit cells. For even  $L_z$  a finite localization length is apparent, which however increases with  $L_z$ . It therefore appears that for large  $L_z$  the even curves will converge to the odd curves, meaning a lack of localization for  $L_z \gg 1$ .

We note that a similar behavior was obtained in Ref.<sup>38</sup> for symplectic multichannel 1D wires. This model is close to ours, but with one important difference. In multichannel 1D wires all the channels are coupled, while in the 2D surface of WTI only nearby channels are coupled.

## Appendix B: The first order quantum corrections to the conductivity

In this appendix we derive the lowest order quantum corrections to the electrical conductivity of WTI and spinless graphene. While the former is our main interest, we find it instructive to compare it to the latter. Our starting point is a low energy effective Hamiltonian for both systems. The Hamiltonian is composed of two decoupled Dirac cones

$$H_0 = -iv_0(\partial_x I^* \otimes s_x + \partial_y I \otimes s_y), \quad (B1)$$

where  $s_i$  are Pauli matrices associated with the spin (sublattice) index of WTI (graphene), and the matrix  $I^*$  is a Pauli matrix associated with the valley index (cf. Eq. (4) in the main text). For WTI,  $I^*$  denotes either  $I_{2 \times 2}$  or  $\tau_z$ , while for graphene  $I^* = \tau_z$ . The corresponding retarded and advanced Green's functions are given by

$$G_0^{R/A}(\mathbf{k}, E) = \frac{E + v_0 k_x I^* \otimes s_x + v_0 k_y I \otimes s_y}{E_{\pm}^2 - (v_0 k)^2}, \quad (B2)$$

where  $E_{\pm} = \lim_{\eta \rightarrow 0^+} E \pm i\eta$ . Time reversal invariant potential disorder is introduced via the matrix  $V(\mathbf{x})$

$$H = H_0 + V(\mathbf{x}), \quad (B3)$$

$$V(\mathbf{x}) = \sum_l v_l(\mathbf{x}) A_l, \quad (B4)$$

where  $A_l$  are  $4 \times 4$  time-reversal symmetric Hermitian matrices of the form  $\tau_i \otimes s_j$  for  $i, j = 0, x, y, z$ . The  $v_l(\mathbf{x})$  are uncorrelated random functions

$$\langle v_l(\mathbf{x}) v_{l'}(\mathbf{x}') \rangle = w_l \delta_{ll'} \delta(\mathbf{x} - \mathbf{x}'). \quad (B5)$$

As mentioned in the main part of the paper, the time-reversal operator  $T$  is different for spinless graphene and WTI. For spinless graphene, the time-reversal operator switches between the two Dirac points, but does not affect the sublattice. Therefore,  $T_g = \tau_x \otimes IK$ , where  $K$  denotes complex conjugation. On the other hand, for WTI it flips the spins but does not affect the valleys, since the Dirac points are at time-reversal-invariant momenta. Therefore,  $T_W = I \otimes s_y K$ . Consequently,  $T_W^2 = -1$  while  $T_g^2 = 1$ . As argued in the main work using the particle diffusion picture, the signs of the quantum interference correction to the conductivity is expected to be given by  $-T^2$ . Another consequence of the difference in  $T$  is that the  $A_l$  matrices which commute with  $T_g$  are all the combinations of  $(I, \tau_x, \tau_y) \otimes (I, s_x, s_z)$  and  $\tau_z \otimes s_y$ , while the matrices which commute with  $T_W$  are  $I \otimes I, \tau_x \otimes I, \tau_z \otimes I, \tau_y \otimes s_x, \tau_y \otimes s_y, \tau_y \otimes s_z$ .

Due to extra symmetries of  $H_0$ , there are additional anti-unitary operators which commute with  $H_0$ . For example, for  $H_0$  in which  $I^* = I$ , all the  $\tau_i T_W$  matrices are such operators. If one chooses disorder that commutes with  $\tau_i T_W$ , rather than with  $T_W$ , then the sign of the quantum correction will be  $-(\tau_i T_W)^2$ .

Our goal is to find the changes in the disorder-averaged conductance as a function of the linear size of the system. The zero-temperature mean longitudinal conductance in the  $x$  direction is given by<sup>30</sup>

$$\sigma_{xx} = \frac{e^2}{2\pi\hbar} \left\langle \int \frac{d^2p}{(2\pi)^2} \text{Tr}[J_x G^R(\mathbf{p}, E_F) J_x G^A(\mathbf{p}, E_F)] \right\rangle,$$

where  $\langle \dots \rangle$  denotes averaging over disorder, and  $J_x = v_0 I^* \otimes s_x$  is the current operator. The diagrammatic way to find this mean value combines the disorder averaged Green's function, the vertex correction, the Cooperon and the dressed Hikami box.

Our derivation follows McCann *et al.* in Ref.<sup>30</sup> for spinless graphene, but with three substantial differences. First, we address here both the WTI and graphene simultaneously in a way that emphasizes the differences between them, both in the Hamiltonian and in the resulting correction. Second, the only assumption we make on the disorder is that it is symmetric with respect to time reversal. We do not assume a dominance of one type of scattering over another. Consequently, the numerical prefactor of the  $\beta$  function depends on the details of the disorder, and these details may affect it by a factor of up to 1/3. Last, since the spectrum of WTI far from the Dirac point is not universal, we adopt a different regularization approach for diverging integrals. Instead of introducing a triangular wrapping, we limit the minimal length scale of the scatterers. For alternative approaches for dealing with this issue see Refs.<sup>29,39</sup>.

We begin with calculating the self-energy within the self consistent Born approximation, given by

$$\Sigma_1^R(\mathbf{q}, E) = \sum_l w_l \int \frac{d^2p}{(2\pi)^2} A_l G^R A_l. \quad (\text{B6})$$

Since  $A_l^2 = I \otimes I$ , and the angular integration over  $\mathbf{p}$  leaves only the diagonal term in  $G^R$ ,

$$[\Sigma_1^R]_{ij}(\mathbf{q}, E) = \delta_{ij} \left[ i\Gamma + \frac{2\Gamma}{\pi} \ln\left(\frac{v_0\Lambda}{E}\right) \right]. \quad (\text{B7})$$

where  $i, j = 1..4$ . In the limit of weak disorder

$$\Gamma = \frac{(\sum_l w_l)E}{4v_0^2}. \quad (\text{B8})$$

The level width  $\Gamma$  is related to the mean free time  $\tau$  through  $\tau = 1/2\Gamma$ . In order to obtain Eq. (B8) we have introduced an ultra violet cutoff  $\Lambda$ , which physically corresponds to the characteristic inverse size of the impurities, and we assumed that  $v_0\Lambda$  is much smaller than the

bulk gap (in WTI) or the bandwidth (in graphene). In the following we ignore the real part of the self-energy, since it only corresponds to a shift in the energy. The disorder averaged Green's function is now given by

$$G^{R/A}(k, E) \approx \frac{E \pm i\Gamma + v_0 k_x I^* \otimes s_x + v_0 k_y I \otimes s_y}{(E \pm i\Gamma)^2 - (v_0 k)^2}.$$

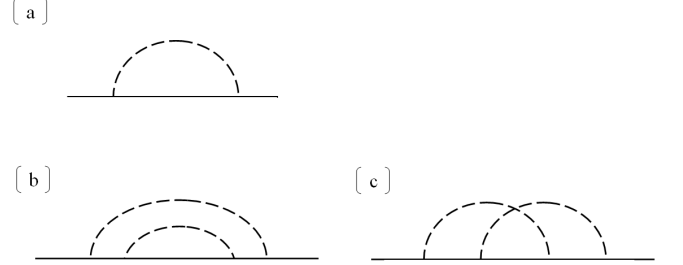


FIG. 5: The leading self-energy diagrams in the Born approximation: (a) first order, and (b), (c) second order.

The self-consistent Born approximation includes diagrams where disorder lines do not intersect, such as depicted in Figs. 5(a) and 5(b). It leaves out diagrams where disorder lines intersect, such as Fig. 5(c). We find the self-consistent Born approximation to be valid when

$$\frac{\Gamma}{E} \ll 1 \quad \text{and} \quad \alpha \equiv \frac{\Gamma}{E} \ln \frac{v_0\Lambda}{E} \ll 1. \quad (\text{B9})$$

Note that due to the logarithmic factor, even when  $v_0\Lambda$  becomes much larger than  $E$  there is still a wide parameter range in which the conditions are satisfied. Moreover, within this range the omission of the real part of the self energy is consistent.

In the limit of weak disorder diagram of Fig. 5(b) is included in our approximation, but its contribution is smaller by a factor of  $\alpha$  relative to that of Fig. 5(a). Its contribution is

$$\Sigma_{2a}^R(\mathbf{q}, E) = \sum_{l, l'} w_l w_{l'} \int \frac{d^2p_1 d^2p_2}{(2\pi)^4} \times \quad (\text{B10})$$

$$A_l G_0^R(\mathbf{p}_1, E) A_{l'} G_0^R(\mathbf{p}_1 - \mathbf{p}_2, E) A_l G_0^R(\mathbf{p}_1, E) A_{l'}.$$

Due to the nested structure of the diagram, the integration over the two loops can be carried separately. The contribution of the diagram is therefore

$$\Sigma_{2a}^R \sim (\Gamma/E_F)^2 \ln^2(v_0\Lambda/E_F) \sim \alpha^2. \quad (\text{B11})$$

Indeed this contribution is negligible for  $\alpha \ll 1$ . The crossed diagram, which is depicted in Fig. 5(c), can also be shown to be of  $O(\alpha^2)$ .

The self-consistent equation of the vertex correction is schematically illustrated in Fig. 6(a). If we denote the



corrected vertex by  $\bar{J}_x$ , then

$$\bar{J}_x(\mathbf{q}, E) = v_0 I^* \otimes s_x \quad (\text{B12})$$

$$+ \sum_l w_l \int \frac{d^2 p}{(2\pi)^2} A_l G^R(\mathbf{p}, E) \bar{J}_x(\mathbf{q}, E) G^A(\mathbf{p} + \mathbf{q}, E) A_l.$$

For  $\mathbf{q} = 0$  we guess a solution of the form  $\bar{J}_x(0, E) = f v_0 I^* \otimes s_x$ , which gives

$$f I^* \otimes s_x = I^* \otimes s_x + f \sum_l w_l A_l (I^* \otimes s_x) A_l \quad (\text{B13})$$

$$\times \int \frac{d^2 p}{(2\pi)^2} \frac{E^2 + \Gamma^2 + v_0^2 p_x^2 - v_0^2 p_y^2}{[(E + i\Gamma)^2 - v_0^2 p^2][(E - i\Gamma)^2 - v_0^2 p^2]}.$$

Due to  $x$ - $y$  symmetry the terms with momenta in the numerator vanish. Moreover,  $A_l(I^* \otimes s_x) = \xi_l(I^* \otimes s_x)A_l$ , where  $\xi_l = \pm 1$ . Therefore  $\sum_l w_l A_l(I^* \otimes s_x)A_l = (\sum_l \xi_l w_l)(I^* \otimes s_x)$ , and the matrix structure of the equation is satisfied. After integrating we find that

$$f = \left(1 - \frac{1}{2} \frac{\sum_l \xi_l w_l}{\sum_l w_l}\right)^{-1}, \quad (\text{B14})$$

where we used the fact that  $\Gamma \ll E$ . We can therefore conclude that the vertex correction is

$$\frac{2}{3} \leq f \leq 2. \quad (\text{B15})$$

The next task is to solve the self-consistent equation of the Cooperon, which is depicted in Fig. 6(b),

$$C_{(ij)(nm)}(\mathbf{k}, \mathbf{k}''; E, \mathbf{Q}, \omega) = \quad (\text{B16})$$

$$\int \frac{d^2 k'}{(2\pi)^2} V_{(ij)(i'j')} \Pi_{(i'j')(i''j'')}(\mathbf{k}'; E, \mathbf{Q}, \omega) V_{(i''j'')(nm)}$$

$$+ \int \frac{d^2 k'}{(2\pi)^2} V_{(ij)(i'j')} \Pi_{(i'j')(i''j'')}(\mathbf{k}'; E, \mathbf{Q}, \omega)$$

$$\times C_{(i''j'')(nm)}(\mathbf{k}', \mathbf{k}''; E, \mathbf{Q}, \omega),$$

$$V_{(ij)(nm)} = \sum_l w_l [A_l]_{in} [A_l]_{jm}, \quad (\text{B17})$$

$$\Pi_{(ij)(nm)}(\mathbf{k}'; E, \mathbf{Q}, \omega) = \quad (\text{B18})$$

$$[G^R(\mathbf{k}' + \mathbf{Q}, E + \omega)]_{in} [G^A(-\mathbf{k}', E)]_{jm}.$$

This equation can be considered as a matrix equation for  $C(E, \mathbf{Q}, \omega)$ , which acts on the vector space  $|\mathbf{k}\rangle \otimes |ij\rangle$ , where  $\mathbf{k}$  denotes the momenta, and  $ij$  denote the internal degrees of freedom of the two particles (of dimension 16).

Anticipating an infrared divergence which is proportional to a diffusive propagator, the Cooperon may be presented as

$$C_{(ij)(nm)}(\mathbf{k}, \mathbf{k}''; E, \mathbf{Q}, \omega) = c \frac{|d\rangle \langle d|}{DQ^2 - i\omega} \quad (\text{B19})$$

$$+ (\text{regular terms}).$$

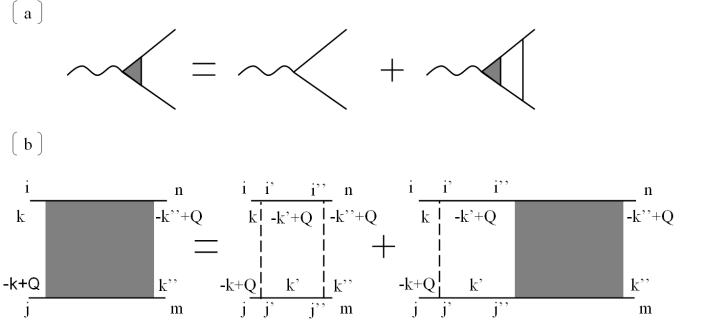


FIG. 6: Diagrammatic representation of the self-consistent equation for (a) the dressed vertex and (b) the Cooperon.

Plugging this ansatz into Eq. (B16), we can extract an equation for the diverging term

$$c(1 - V\Pi) \frac{|d\rangle \langle d|}{DQ^2 - i\omega} = V\Pi V. \quad (\text{B20})$$

Multiplying from the left with  $(V\Pi V)^{-1}$  and from the right with  $|d\rangle$ , gives an eigenstate equation for the diffusive mode

$$(\Pi V)^{-1} (V^{-1} - \Pi) |d\rangle = c^{-1} (DQ^2 - i\omega) |d\rangle, \quad (\text{B21})$$

where

$$(V^{-1} - \Pi) = \left( \sum_l w_l A_l \otimes A_l \right)^{-1} \quad (\text{B22})$$

$$- \int \frac{d^2 k}{(2\pi)^2} \frac{1}{[(E + \omega + i\Gamma)^2 - v_0^2 k^2][(E - i\Gamma)^2 - v_0^2 k^2]}$$

$$\times (E + \omega + i\Gamma + v_0 k_x I^* \otimes s_x + v_0 k_y I \otimes s_y)$$

$$\otimes (E - i\Gamma - v_0 k_x I^* \otimes s_x - v_0 k_y I \otimes s_y).$$

The terms which are linear in  $k_x$  and  $k_y$  vanish in the integration, and the remaining three integrals, which multiply the three matrices  $(I \otimes I) \otimes (I \otimes I)$ ,  $(I^* \otimes s_x) \otimes (I^* \otimes s_x)$  and  $(I \otimes s_y) \otimes (I \otimes s_y)$ , are respectively

$$\int \frac{d^2 k}{(2\pi)^2} \frac{(E + \omega + i\Gamma)(E - i\Gamma)}{[(E + i\Gamma)^2 - v_0^2 k^2][(E - i\Gamma)^2 - v_0^2 k^2]} \quad (\text{B23})$$

$$= - \frac{(\sum_l w_l)^{-1}}{2},$$

$$\int \frac{d^2 k}{(2\pi)^2} \frac{-v_0^2 k_x^2}{[(E + i\Gamma)^2 - v_0^2 k^2][(E - i\Gamma)^2 - v_0^2 k^2]} \quad (\text{B24})$$

$$= \frac{(\sum_l w_l)^{-1}}{4} + O(\Gamma/E, \alpha),$$

$$\int \frac{d^2 k}{(2\pi)^2} \frac{-v_0^2 k_y^2}{[(E + i\Gamma)^2 - v_0^2 k^2][(E - i\Gamma)^2 - v_0^2 k^2]} = \quad (\text{B25})$$

$$= \frac{(\sum_l w_l)^{-1}}{4} + O(\Gamma/E, \alpha).$$

Therefore

$$(V^{-1} - \Pi) \approx \left( \sum_l w_l A_l \otimes A_l \right)^{-1} - \left( \sum_l w_l \right)^{-1} \quad (\text{B26})$$

$$\times \frac{1}{2} \left( 1 - \frac{1}{2} [(I^* \otimes s_x) \otimes (I^* \otimes s_x) + (I \otimes s_y) \otimes (I \otimes s_y)] \right).$$

We are interested in the zero mode of the above matrix. Since  $V$  mixes all momenta equally, any eigenvector of  $V^{-1}$  which depends on momenta will have a diverging eigenvalue, and cannot give rise to a zero mode in the above equation. For generic disorder which respects time-reversal symmetry we find that the zero mode for WTI (with  $I^* = I$  for concreteness) and graphene are given by

$$\langle \mathbf{k}, ij | d_W \rangle = \delta_{\tau_i, \tau_j} (\delta_{s_i, 1} \delta_{s_j, -1} - \delta_{s_i, -1} \delta_{s_j, 1}) / 2,$$

$$\langle \mathbf{k}, ij | d_g \rangle = \delta_{s_i, s_j} (\delta_{\tau_i, 1} \delta_{\tau_j, -1} + \delta_{\tau_i, -1} \delta_{\tau_j, 1}) / 2,$$

where  $\tau_i$  ( $s_i$ ) is the valley (spin/pseudospin) subindex of the index  $i$ . This can be easily verified by the facts that  $A_l |d\rangle = |d\rangle$ ,  $(I^* \otimes s_x) \otimes (I^* \otimes s_x) |d\rangle = -|d\rangle$ , and  $(I \otimes s_y) \otimes (I \otimes s_y) |d\rangle = -|d\rangle$ . Note that the vector  $|d\rangle$  has an eigenvalue of 1 with respect to  $\Pi V$ , and therefore  $(\Pi V)^{-1} |d\rangle = |d\rangle$ .

The diffusion coefficient  $D$  and the constant  $c$  from Eq. (B19) can be extracted by expanding Eq. (B21) in  $\omega$  and  $\mathbf{Q}$ . After some algebra one finds that in both cases

$$c = 8v_0^2 \frac{\Gamma^2}{E}, \quad (\text{B27})$$

$$D = \frac{v_0^2}{2\Gamma}. \quad (\text{B28})$$

Note that we keep  $c$  although it is of lower order in  $\Gamma/E$ , since it is associated with the divergence of the Cooperon.

The leading term of the conductivity  $\sigma_{xx}$  is given by

$$\sigma_{xx}^0 = \frac{e^2}{2\pi\hbar} \int \frac{d^2 p}{(2\pi)^2} \text{Tr}[\tilde{J}_x G^R(\mathbf{p}, E_F) J_x G^A(\mathbf{p}, E_F)]$$

$$= \frac{e^2}{\hbar} \frac{E}{\pi v_0^2} D \frac{f_v}{2} = \frac{e^2}{\hbar} \frac{f_v}{2} \frac{E}{\Gamma}. \quad (\text{B29})$$

The first quantum interference correction  $\delta\sigma_{xx}^a$ , which is depicted in Fig. 7(a), is given by

$$\delta\sigma_{xx}^a = \frac{e^2}{2\pi\hbar} \int \frac{d^2 k d^2 Q}{(2\pi)^4} [\bar{J}_x]_{i'i} [\bar{J}_x]_{jj'} G_{j'i_2'}^A(\mathbf{k}, E_F) \quad (\text{B30})$$

$$\times G_{i_1 i'}^A(-\mathbf{k} + \mathbf{Q}, E_F) G_{ii_1}^R(-\mathbf{k} + \mathbf{Q}, E_F)$$

$$\times G_{i_2 j}^R(\mathbf{k}, E_F) C_{(i_1 i_2)(i_2 i_1')}(\mathbf{k}, \mathbf{k}; E_F, \mathbf{Q}, 0).$$

The divergent contribution to the correction comes from the limit of  $\mathbf{Q} = 0$  in the Green's functions, where

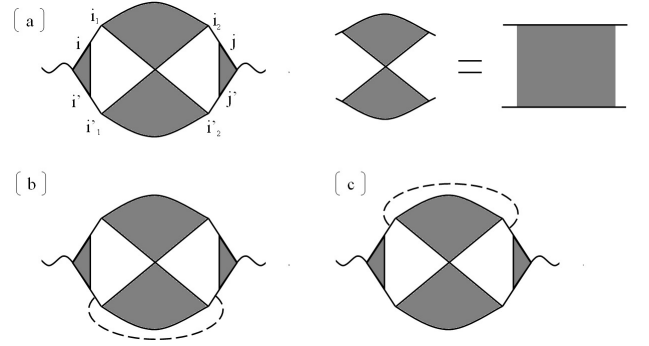


FIG. 7: The leading order quantum corrections to the conductivity. These diagrams can be viewed as a combination of a dressed Hikami box with the Cooperon.

they are regular. Henceforth

$$\delta\sigma_{xx}^a \approx \frac{e^2 v_0^2}{2\pi\hbar} f^2 \int \frac{d^2 Q}{(2\pi)^2} \frac{c}{D Q^2} \int \frac{d^2 k}{(2\pi)^2} \quad (\text{B31})$$

$$\times [I^* \otimes s_x]_{i'i} [I^* \otimes s_x]_{jj'} G_{j'i_2'}^A(\mathbf{k}) G_{i_1 i'}^A(-\mathbf{k})$$

$$\times G_{ii_1}^R(-\mathbf{k}) G_{i_2 j}^R(\mathbf{k}) \langle i_1 i_2 | d \rangle \langle i_2, i_1' | d \rangle.$$

Keeping only the divergent part of the integral over  $\mathbf{Q}$ , and noticing that the index summation is actually a trace, we have

$$\delta\sigma_{xx}^a = \ln(L) \frac{e^2 v_0^2}{4\pi^2 \hbar} \frac{c}{D} f^2 \int \frac{d^2 k}{(2\pi)^2} \quad (\text{B32})$$

$$\times \text{Tr} [G^A(-\mathbf{k}) (I^* \otimes s_x) G^R(-\mathbf{k}) |d\rangle$$

$$\times G^A(\mathbf{k})^T (I^* \otimes s_x)^T G^R(\mathbf{k})^T |d\rangle],$$

where the matrix  $|d\rangle$  is defined by  $[|d\rangle]_{ij} = \langle ij | d \rangle$ . Using the fact that

$$G^A(-\mathbf{k}) (I^* \otimes s_x) G^R(-\mathbf{k}) \propto 2v_0^2 k_x k_y (I \otimes s_y)$$

$$+ (E^2 + \Gamma^2 + v_0^2 k_x^2 - v_0^2 k_y^2) (I^* \otimes s_x)$$

$$- 2Ev_0 k_x (I \otimes I) - 2v_0 k_y \Gamma (I^* \otimes s_z), \quad (\text{B33})$$

the only non vanishing and non negligible traces are those which are proportional to  $\text{Tr}[(I^* \otimes s_x) |d\rangle (I^* \otimes s_x) |d\rangle]$ ,  $\text{Tr}[|d\rangle (I \otimes s_y) |d\rangle (I \otimes s_y)]$ , and  $\text{Tr}[|d\rangle |d\rangle]$ . The resulting

correction is now

$$\begin{aligned}
\delta\sigma_{xx}^a &\approx -\text{Tr}[[d]^2]\ln(L)\frac{e^2v_0^2}{4\pi^2\hbar}\frac{c}{D}f^2\times \\
&\int\frac{d^2k}{(2\pi)^2}\frac{E^4+v_0^4k^4+4E^2v_0^2k_x^2}{[(E+i\Gamma)^2-v_0^2k^2]^2[(E-i\Gamma)^2-v_0^2k^2]^2} \\
&\approx -\text{Tr}[[d]^2]\ln(L)\frac{e^2v_0^2}{2\pi^2\hbar}\frac{c}{D}f^2\times \\
&\int\frac{d^2k}{(2\pi)^2}\frac{4E^4}{[(E+i\Gamma)^2-v_0^2k^2]^2[(E-i\Gamma)^2-v_0^2k^2]^2} \\
&= -\text{Tr}[[d]^2]\ln(L)\frac{e^2v_0^2}{4\pi^2\hbar}\frac{c}{D}f^2\frac{E}{16\Gamma^3} \\
&= -T^2\ln(L)f^2\frac{1}{4\pi^2}\frac{e^2}{\hbar}, \tag{B34}
\end{aligned}$$

where we replaced  $[[d]]^2$  with  $T^2$ , since they are equal.

As first noted by Ref.<sup>30</sup>, the extra quantum corrections to the conductivity, which are depicted in Figs. 7(b,c), are non vanishing due to the independence on momenta of the current vertex. These two contributions are equal, and are given by

$$\delta\sigma_{xx}^b = \delta\sigma_{xx}^c = -\frac{f'}{4}\delta\sigma_{xx}^a, \tag{B35}$$

$$\begin{aligned}
f' &= \frac{1}{4\sum_l w_l} \\
&\times \left\{ \begin{array}{ll} \sum_l w_l \text{Tr}[A_l(I \otimes s_z)A_l^T(I \otimes s_z)] & \text{WTI} \\ \sum_l w_l \text{Tr}[A_l(\tau_y \otimes s_x)A_l^T(\tau_y \otimes s_x)] & \text{graphene} \end{array} \right. \tag{B36}
\end{aligned}$$

Since  $-1 \leq f' \leq 1$ , the sign of the quantum correction is still determined entirely by  $T^2$ .

We have shown above that  $T_g^2 = 1$  while  $T_W^2 = -1$ . Therefore we can conclude from Eqs. (B34)-(B36) that spinless graphene tends to be localized, while a WTI flows towards perfect conduction.

### Appendix C: Weak localization and the time-reversal operator

In this appendix we provide a straightforward explanation for the fact that the sign of the weak localization correction is the same as the sign of the time-reversal operator squared ( $T^2$ ). To this end, we express the Green's function as a sum over amplitudes associated with trajectories. Similarly, we express the return probability as a sum over products of such amplitudes. The coherent contributions that give rise to weak localization/antilocalization come from products of time-reversal conjugate trajectories. By analyzing the action of  $T$  on trajectories the above relation is established.

Consider the Dyson series for the Green's function  $G$ ,

$$G = G^0 \sum_{n=0}^{\infty} (VG^0)^n, \tag{C1}$$

where  $G^0$  is the clean Green's function, and  $V$  is the disorder potential. The matrix element of  $G$  that connects the lattice site  $i$  and spin state  $\sigma$  with the lattice site  $j$  and spin state  $\sigma'$  may be written as a sum over trajectories that connect these two sites and spin states, and which go through a series of intermediate points  $\alpha = (i\sigma, i_n\sigma_n, i_{n-1}\sigma_{n-1}, \dots, i_1\sigma_1, j\sigma')$

$$G_{i\sigma,j\sigma'} = \sum_{\alpha} \mathcal{A}_{i\sigma,j\sigma'}^{\alpha}, \tag{C2}$$

$$\mathcal{A}_{i\sigma,j\sigma'}^{\alpha} = G_{i\sigma,i_n\sigma_n}^0 \cdot V_{i_n\sigma_n,i_{n-1}\sigma_{n-1}} \cdot \dots \cdot G_{i_1\sigma_1,j\sigma'}^0. \tag{C3}$$

Given that the system is symmetric to some anti-unitary operator, most notably the time-reversal operator  $T$ , we define  $|\bar{\sigma}\rangle = \xi_{\sigma}T|\sigma\rangle$ , where  $\xi_{\sigma} = \pm 1$ . Consequently,  $G_{i\sigma,j\sigma'}^0 = \xi_{\sigma}\xi_{\sigma'}G_{j\bar{\sigma}',i\bar{\sigma}}^0$  and  $V_{i\sigma,j\sigma'} = \xi_{\sigma}\xi_{\sigma'}V_{j\bar{\sigma}',i\bar{\sigma}}$ . A straightforward manipulation then yields

$$\mathcal{A}_{i\sigma,i\sigma'}^{\alpha} = \xi_{\sigma}\xi_{\sigma'}\mathcal{A}_{i\bar{\sigma},i\bar{\sigma}}^{\bar{\alpha}}, \tag{C4}$$

where  $\bar{\alpha} = (i\bar{\sigma}', i_1\bar{\sigma}_1, \dots, i\bar{\sigma})$ . Note that all the sign factors except  $\xi_{\sigma}$  and  $\xi_{\sigma'}$  appear twice, and therefore are canceled out.

Using (C2) we find that the probability of a particle to return back to its initial site, with perhaps a different spin state, is given by

$$|G_{i\sigma,i\sigma'}|^2 = \sum_{\alpha,\alpha'} \mathcal{A}_{i\sigma,i\sigma'}^{\alpha} (\mathcal{A}_{i\sigma,i\sigma'}^{\alpha'})^* \tag{C5}$$

Two types of pairs of trajectories contribute coherently to the disorder-averaged double sum in equation (C5), since their phases do not fluctuate. The obvious contribution is the classical contribution consisting of pairs with  $\alpha = \alpha'$ . However, due to  $T$ -symmetry, an additional contribution exists in which  $\alpha$  comes paired with  $\bar{\alpha}$ . Comparing equation (C4) with equation (C5) one finds that pairs of time conjugated paths may appear only if  $\sigma' = \bar{\sigma}$ . Therefore whenever it appears, the sign factor of such term is  $\xi_{\sigma}\xi_{\bar{\sigma}} = \langle\sigma|T^2|\sigma\rangle = \text{sign}(T^2)$ . Also notice that the size of this term is equal to the size of the classical term, and therefore may either double or suppress it. Hence for  $T^2 = -1(1)$  the probability for a diffusing particle to return to its original position is higher (lower) than the classical probability, and this is an indication for weak anti-localization (weak localization). If the Hamiltonian commutes with more than one anti-unitary operator, for example in the case of a spin independent Hamiltonian, the total correction is composed of the contributions from all the different trajectories with  $\sigma' = \bar{\sigma}$ .

\* Contributed equally to this work.

- <sup>1</sup> M. Z. Hasan and C. L. Kane, Rev. Mod. Phys. **82**, 3045 (2010).
- <sup>2</sup> X.-L. Qi and S.-C. Zhang, Rev. Mod. Phys. **83**, 1057 (2011).
- <sup>3</sup> M. König, S. Wiedmann, C. Brüne, A. Roth, H. Buhmann, L. W. Molenkamp, X.-L. Qi, and S.-C. Zhang, Science **318**, 766 (2007).
- <sup>4</sup> D. Hsieh, D. Qian, L. Wray, Y. Xia, Y. S. Hor, R. J. Cava, and M. Z. Hasan, Nature **452**, 970 (2008).
- <sup>5</sup> C. L. Kane and E. J. Mele, Phys. Rev. Lett. **95**, 146802 (2005).
- <sup>6</sup> L. Fu, C. L. Kane, and E. J. Mele, Phys. Rev. Lett. **98**, 106803 (2007).
- <sup>7</sup> J. E. Moore and L. Balents, Phys. Rev. B **75**, 121306 (2007).
- <sup>8</sup> R. Roy, Phys. Rev. B **79**, 195322 (2009).
- <sup>9</sup> Y. Ran, Y. Zhang, and A. Vishwanath, Nature Phys. **5**, 298 (2009).
- <sup>10</sup> J. H. Bardarson, J. Tworzydło, P. W. Brouwer, and C. W. J. Beenakker, Phys. Rev. Lett. **99**, 106801 (2007).
- <sup>11</sup> K. Nomura, M. Koshino, and S. Ryu, Phys. Rev. Lett. **99**, 146806 (2007).
- <sup>12</sup> P. M. Ostrovsky, I. V. Gornyi, and A. D. Mirlin, Eur. Phys. J. Special Topics **148**, 63 (2007).
- <sup>13</sup> S. Ryu, C. Mudry, H. Obuse, and A. Furusaki, Phys. Rev. Lett. **99**, 116601 (2007).
- <sup>14</sup> A. D. Mirlin, F. Evers, I. V. Gornyi, and P. M. Ostrovsky, *50 years of Anderson localization* (World Scientific, Singapore, 2010), p. 107.
- <sup>15</sup> L. Fu and C. L. Kane, Phys. Rev. B **76**, 045302 (2007).
- <sup>16</sup> J. von Neumann and E. P. Wigner, Z. Phys **30**, 467 (1929).
- <sup>17</sup> A. M. Essin and J. E. Moore, Phys. Rev. B **76**, 165307 (2007).
- <sup>18</sup> Z. Ringel and Y. E. Kraus, Phys. Rev. B **83**, 245115 (2011).
- <sup>19</sup> F. Bloch, Phys. Rev. B **2**, 109 (1970).
- <sup>20</sup> D. J. Thouless, Physics Reports **13**, 93 (1974).
- <sup>21</sup> E. Abrahams, P. W. Anderson, D. C. Licciardello, and T. V. Ramakrishnan, Phys. Rev. Lett. **42**, 673 (1979).
- <sup>22</sup> B. T. Debney, Journal of Physics C: Solid State Physics **10**, 4719 (1977).
- <sup>23</sup> W. Kohn, Phys. Rev. **133**, A171 (1964).
- <sup>24</sup> P. A. Lee and T. V. Ramakrishnan, Rev. Mod. Phys. **57**, 287 (1985).
- <sup>25</sup> P. W. Anderson and P. A. Lee, Prog. Theor. Phys. Supplement **69**, 212 (1980).
- <sup>26</sup> E. Akkermans and G. Montambaux, Phys. Rev. Lett. **68**, 642 (1992).
- <sup>27</sup> B. L. Altshuler and A. Aronov, *Electron-Electron Interactions in Disordered Systems* (North-Holland, Amsterdam, 1985), p. 27.
- <sup>28</sup> This correction is of zeroth order in  $\ln(\Delta/E)/(E\tau)$ . For a first order corrections see Ref.<sup>29</sup>.
- <sup>29</sup> I. L. Aleiner and K. B. Efetov, Phys. Rev. Lett. **97**, 236801 (2006).
- <sup>30</sup> E. McCann, K. Kechedzhi, V. I. Fal'ko, H. Suzuura, T. Ando, and B. L. Altshuler, Phys. Rev. Lett. **97**, 146805 (2006).
- <sup>31</sup> S. Murakami, S. Iso, Y. Avishai, M. Onoda, and N. Nagaosa, Phys. Rev. B **76**, 205304 (2007).
- <sup>32</sup> M. Onoda, Y. Avishai, and N. Nagaosa, Phys. Rev. Lett. **98**, 076802 (2007).
- <sup>33</sup> M. B. H. T. A. Loring, EPL **76**, 67004 (2010).
- <sup>34</sup> H. Obuse, A. Furusaki, S. Ryu, and C. Mudry, Phys. Rev. B **76**, 075301 (2007).
- <sup>35</sup> R. S. K. Mong, J. H. Bardarson, and J. E. Moore, Phys. Rev. Lett. **108**, 076804 (2012).
- <sup>36</sup> A. Auerbach, *Interacting Electrons and Quantum Magnetism* (Springer-Verlag, New York, 1994).
- <sup>37</sup> Ref.<sup>35</sup> attempts to find the RG flow of the conductance by studying numerically the low energy Hamiltonian, Eq. (4). This study finds that for any disorder strength the flow is towards a perfect metal. It is known however, see e.g. Ref.<sup>29</sup>, that for intermediate and strong disorder in Dirac Hamiltonians, ultra-violet cut-off effects become important and may alter the conductance considerably. In particular, different cut-offs reflect different topological properties of the 3D bulk, most notably the pairs switching behavior. We believe that the stability of the numerical results to the choice of cut-offs should be investigated before the RG flow of the conductance may be concluded from numerical calculations.
- <sup>38</sup> Y. Takane, J. Phys. Soc. Jpn. **73**, 2366 (2004).
- <sup>39</sup> P. M. Ostrovsky, I. V. Gornyi, and A. D. Mirlin, Phys. Rev. B **74**, 235443 (2006).